APS physics

CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Granger causality analysis with nonuniform sampling and its application to pulse-coupled nonlinear dynamics

Yaoyu Zhang, Yanyang Xiao, Douglas Zhou, and David Cai

Phys. Rev. E **93**, 042217 — Published 26 April 2016

DOI: 10.1103/PhysRevE.93.042217

Nonuniform Sampling Granger Causality Analysis and Its Application to Pulse-coupled Nonlinear Dynamics

Yaoyu Zhang,[1] Yanyang Xiao,[1] Douglas Zhou,[1,*] and David Cai[1,2,3,†]

[1] *Department of Mathematics, MOE-LSC, and Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, China*

[2] *Courant Institute of Mathematical Sciences and Center for Neural Science, New York University, New York, NY, USA*

[3] *NYUAD Institute, New York University Abu Dhabi, Abu Dhabi, UAE*

(Dated: March 14, 2016)

The Granger causality (GC) analysis is an effective approach to infer causal relations for time series. However, for data obtained by uniform sampling (*i.e.*, with equal sampling time interval), it is known that GC can yield unreliable causal inference due to aliasing if the sampling rate is not sufficiently high. To solve this unreliability issue, we consider the nonuniform sampling scheme as it can mitigate against aliasing. By developing an unbiased estimation of power spectral density of nonuniformly sampled time series, we establish a framework of spectrum-based nonparametric GC analysis. Applying this framework to a general class of pulse-coupled nonlinear networks and utilizing some particular spectral structure possessed by these nonlinear network data, we demonstrate that, for such nonlinear networks with nonuniformly sampled data, reliable GC inference can be achieved at low nonuniform mean sampling rate at which the traditional uniform sampling GC may lead to spurious causal inference.

PACS numbers: 05.45.Tp, 02.50.Tt, 05.10.-a, 84.35.+i, 07.05.Kf

I. INTRODUCTION

The Granger causality (GC) analysis is an important approach in the detection of causal relations among time series [1, 2]. Based on the idea that the driver is always earlier than the recipient, GC is defined as the reduction of the prediction error of one time series by incorporating the history information of the other time series through a joint-regression. Due to its intuitive conceptual appeal and easy implementation, GC has been widely applied in various scientific fields, *e.g.*, physics [3, 4], biology [5–7], neuroscience [8–10], economics [11–13], and social sciences [14, 15].

In network sciences, GC analysis based on linear regression/prediction has emerged as a powerful methodology for network topology reconstruction as it can often establish structural connections through directional causal interactions among different nodes. The GC theory is well established for linear systems. However, whether it can be successfully applied to nonlinear systems is still under active investigations. Recently, for multivariate nonlinear network data from functional magnetic resonance imaging (fMRI), it has been found that GC can reflect the underlying vascular anatomic structure [16]. For a general class of pulse-coupled nonlinear networks which arises in the modeling of neuronal dynamics, it has been demonstrated that the structural connectivity can be successfully reconstructed through the GC analysis [17, 18].

In applications, it is critical to assess the reliability of GC inference [19]. The original GC analysis is based on linear-regression with discrete time series. However, physical quantities of real systems are usually continuous in time. One needs to address the important issue of whether the reliability of GC inference is contingent upon different sampling strategies. As pointed out in Ref. [20] for uniformly sampled time series, a key factor in determining the GC reliability is the sampling interval length τ. Following Ref. [20], we will term GC value as a function of τ the GC sampling structure. For a broad class of uniformly sampled time series, *i.e.*, with equal sampling interval, it has been found that, if the sampling interval is not sufficiently fine, the GC sampling structure displays oscillatory decay features. These features can give rise to various types of GC inference hazards [20, 21]. Furthermore, for the case where there exists causal influence, the GC value tends to 0 linearly as τ approaches 0 despite the fact that more information is incorporated. This scaling behavior presents a paradox because it seems not consistent with the intuition that, as one samples finer and finer to include more high frequency information for GC analysis, one would expect that GC should yield a nonzero value if the causal influence exists. Obviously, for uniformly sampled time series, the above phenomena in GC analysis greatly complicate the GC inference. In the framework of uniform sampling, these inference hazards can be removed by using a normalized GC with a sufficiently fine sampling [20, 21]. However, for experimental recordings with a relatively low sampling rate, the above method is rendered incapable of providing consistent GC inference.

On the other hand, nonuniform sampling schemes have been extensively studied [22–24]. The significance of

* zdz@sjtu.edu.cn

† cai@cims.nyu.edu

such schemes is twofold: (i) In research fields such as Astronomy [25–27], Seismology [28–30], Paleoclimatology [31, 32], Genetics [33, 34], and Biomedical Imaging [35, 36], certain quantities can only be observed in a nonuniform manner due to spatiotemporal constraint in signal recordings; (ii) In contrast to uniform sampling, a nonuniform sampling allows for low sampling rate but still possessing reliable spectral information of signals. For uniform sampling, as is well known, there is the Nyquist rate below which the original signal and its spectrum cannot be reconstructed reliably (*i.e.*, aliasing). However, the nonuniform sampling strategy enables one to sample signals with various sizes of sampling intervals, thus preserving both low and high frequency information of signals while the nonuniform mean sampling rate can still be lower than the Nyquist rate [22, 37, 38]. Here, the nonuniform sampling interval length τ is defined as the mean length of nonuniform sampling intervals, thereby, $1/\tau$ is the corresponding nonuniform mean sampling rate. Due to their anti-aliasing advantages, nonuniform sampling schemes are sometimes intentionally designed with certain random sampling devices for efficient and accurate data processing in applications [24, 39–41].

In GC analysis with uniformly sampled time series, which will be referred to as the uniform sampling GC, as the sampling interval length τ increases, not only high frequency spectral information is lost, but the low frequency spectral component is also corrupted due to aliasing. The aliasing behavior is closely related to the unreliability of the uniform sampling GC analysis. A nonuniform sampling scheme appears to be a good candidate to remove the sampling artifacts to achieve a reliable GC analysis. However, the conventional GC analysis cannot be naturally extended to nonuniformly sampled time series because it is based on linear-regression, in which one needs to fit regression coefficients at each discrete time with an equal lag, *i.e.*, $t-\tau$, $t-2\tau$, $\cdots$, where τ is the fixed sampling interval length. However, for nonuniform sampling, there is no fixed sampling interval length, namely, there is a different time lag at a different time. In this work, we establish a framework of GC analysis for nonuniformly sampled data, which will be referred to as nonuniform sampling GC. First, instead of using the time domain regression, we employ the frequency-domain nonparametric GC theory, which is based on estimation of the power spectral density (PSD) of time series. Note that the spectrum-based nonparametric GC analysis for uniformly sampled time series has already been well established [2, 42]. However, for nonuniformly sampled time series, it has not been formulated and the key point is how to obtain good estimate of PSD. Second, because a direct application of the nonuniform fast Fourier transform (NUFFT) to the PSD estimation of nonuniformly sampled time series can lead to large biases in the estimated PSDs, thus inducing large errors in the GC analysis, we develop an unbiased method for PSD estimation of nonuniformly sampled time series. Third, during the computation of PSD estimation, one has to choose a cutoff frequency $f_{\rm cut}$. However, the nonuniform sampling GC inference varies with different cutoff frequencies $f_{\rm cut}$ of the estimated PSDs even when $f_{\rm cut}$ is sufficiently large to well capture spectral information of signals. We demonstrate that its origin is similar to the scaling paradox in the uniform sampling GC. By using a similar rescaling as in Ref. [20], this issue can be resolved to obtain reliable GC inference.

Using our framework, we extend the conventional uniform sampling GC analysis to the case of nonuniformly sampled data. However, this framework is nonparametric and does not take into account specific structures of the data. One can take advantage of specific structures embedded in the time series to further reduce the PSD estimation error. In this work, we focus on time series obtained from a general class of pulse-coupled nonlinear networks, *i.e.*, integrate-and-fire (I&F) networks, and exploit the structures of their PSDs to achieve a more reliable GC analysis. Although GC analysis is established for linear dynamics, for the pulse-coupled nonlinear networks, it has been found that GC analysis can well capture the underlying structural connectivity [17, 18]. However, directly applying the nonparametric framework to time series obtained from the pulse-coupled nonlinear network dynamics, we find that the estimated PSDs have strong statistical fluctuations, which can significantly corrupt the GC analysis. To overcome this problem, we introduce a covariance truncation method to better estimate PSD based on the exponential decay feature in the covariance of the I&F network dynamics. Note that such feature in the covariance can be observed for a wide class of dynamics with finite memory as discussed in detail in Section IV B. Therefore, our covariance truncation scheme, in general, can also be applied to these dynamics to obtain a more accurate PSD estimation. Furthermore, we demonstrate that the PSD functions of I&F network dynamics possess a power-law decay of specific exponents at high frequencies. Utilizing this decay structure, we develop a parametric tail-fitting scheme to further reduce the fluctuations in PSD estimation, thus leading to a reliable nonuniform sampling GC inference. Note that the power-law decay in the tail of PSD can be generally observed in dynamical systems. By choosing appropriate decay order, our power-law tail-fitting approach thus can be extended to suppress statistical fluctuations in the tail of PSD in such systems as further discussed in Section IV C. We demonstrate that, once the PSD has been well estimated, the remaining procedures for GC analysis are the same as those established for uniformly sampled time series. Therefore, in this paper, we focus more on the PSD estimation in an attempt to develop a spectrum-based nonparametric GC analysis based on nonuniformly sampled time series.

Finally, we present our numerical results for the nonlinear I&F network dynamics and demonstrate that (i) the nonuniform sampling GC can indeed resolve the sampling hazards that may arise in the uniform sampling GC analysis; (ii) reliable inference can be achieved at a

nonuniform mean sampling rate lower than the Nyquist rate; (iii) taking into account the specific structures in the I&F dynamics, our spectral processing procedures can greatly improve the accuracy of PSD estimation and give rise to a reliable nonuniform sampling GC inference; (iv) the inferred GC relations through our nonuniform sampling GC are coincident with the underlying structural connectivity of nonlinear I&F networks.

The article is organized as follows. In Section II, we briefly introduce the pulse-coupled nonlinear system (I&F networks), the conventional GC analysis with uniformly sampled data, and related sampling artifacts in such GC analysis. In Section III, we establish a nonparametric framework for the nonuniform sampling GC analysis. In Section IV, we analyze the PSD properties for time series obtained from I&F networks and describe the corresponding truncation and high-frequency tail-fitting procedure. In Section V, numerical results are presented to demonstrate the validity and reliability of our nonuniform sampling GC analysis. We show that the sampling artifacts occurring in the uniform sampling GC can indeed be removed through our nonuniform sampling GC analysis. In Section VI, we present conclusions and discussions. In Appendix, we provide detailed theoretical derivations necessary for the discussions in the main text.

II. BACKGROUND AND NOTATIONS

A. Pulse-coupled nonlinear systems

The pulse-coupled nonlinear system we consider is an integrate-and-fire (I&F) network with N excitatory nodes. The dynamics of its ith node is described by

$$\begin{aligned}\frac{\mathrm{d}x_i}{\mathrm{d}t} &= -\frac{x_i}{\tau_\mathrm{d}} - g_i\left(x_i - x^\mathrm{E}\right),\\ \frac{\mathrm{d}g_i}{\mathrm{d}t} &= -\frac{g_i}{\sigma} + \sum_{j\neq i}^{N}\sum_k s_{ji}\delta\left(t - T_{j,k}\right) + \lambda\sum_l \delta\left(t - T^\mathrm{P}_{i,l}\right),\end{aligned} \tag{1}$$

where x_i is the state variable of node i, τ_d is the decay time scale, and x^E is the reversal value of excitation. g_i is the input that decays with time constant σ and rises instantaneously at the moments $T_{j,k}$, $T^\mathrm{P}_{i,l}$ with magnitude s_{ji}, λ, respectively. When x_i is less than the firing threshold x^th, the ith node evolves according to Eq. (1). When x_i reaches x^th, it is reset to a reset value x^r and stays for a refractory period τ_ref. The above system arises from many research fields such as image processing, speech recognition, gene regulatory modeling and neuronal dynamics [43–46]. Following the terminology in neuroscience, the moment when x_i reaches x^th is called a spiking event (say, the kth spike of the ith node is denoted by $T_{i,k}$). We will refer to a node as a neuron, x_i as the ith neuron's voltage, and g_i as the ith neuron's conductance in the following discussion. We can capture these spikes of neuron i by the spike train time series, which is defined as

$$S^i_T(t) = \sum_k \delta(t - T_{i,k}), \tag{2}$$

where $\delta(t)$ is the Dirac delta function. In Eq. (1), x_i receives the lth external input at time $T^\mathrm{P}_{i,l}$, which is modeled by a Poisson process with strength λ and rate μ. The interaction between neurons is described by s_{ji} which represents the connection strength from neuron j to neuron i. Note that in the following we use the parameters $x^\mathrm{th} = 1$, $x^\mathrm{r} = 0$, $x^\mathrm{E} = 14/3$, $\tau_\mathrm{ref} = 2\,\mathrm{ms}$, $\sigma = 2\,\mathrm{ms}$ and $\tau_\mathrm{d} = 20\,\mathrm{ms}$, which are typical values for neuronal dynamics [47–49].

B. GC Analysis

The idea of GC is that if a better prediction, *i.e.*, less prediction error, of one variable X based on its history can be achieved by incorporating the history of the other variable Y, then Y is regarded to influence X in the GC sense [2, 10]. More precisely, for the auto-regression on uniformly sampled time series X_t, Y_t, we obtain

$$\begin{aligned}X_t &= \sum_{j=1} a_{1j}X_{t-j} + \epsilon_{1t},\\ Y_t &= \sum_{j=1} d_{1j}Y_{t-j} + \eta_{1t},\end{aligned}$$

where $\mathrm{var}(\epsilon_{1t}) = \Sigma_1$ and $\mathrm{var}(\eta_{1t}) = \Gamma_1$ quantify the prediction error when only their histories are used individually. When one incorporates the history of the other time series, the joint-regression for X_t and Y_t can be expressed as

$$\begin{aligned}X_t &= \sum_{j=1} a_{2j}X_{t-j} + \sum_{j=1} b_{2j}Y_{t-j} + \epsilon_{2t},\\ Y_t &= \sum_{j=1} c_{2j}X_{t-j} + \sum_{j=1} d_{2j}Y_{t-j} + \eta_{2t},\end{aligned}$$

where the variances $\Sigma_2 = \mathrm{var}(\epsilon_{2t})$, $\Gamma_2 = \mathrm{var}(\eta_{2t})$, and the covariance $\Upsilon_2 = \mathrm{cov}\left(\epsilon_{2t},\, \eta_{2t}\right)$ quantify the performance of prediction after the joint-regression. Because one typically cannot obtain a worse prediction, *i.e.*, greater prediction error, when incorporating more information, obviously, $\Sigma_2 \leqslant \Sigma_1$ and $\Gamma_2 \leqslant \Gamma_1$.

The directional GC from Y to X is defined as

$$F_{Y\to X} = \ln\frac{\Sigma_1}{\Sigma_2},$$

whose value conforms with the intuition of causality: a positive value indicates a corresponding directional influence and the zero value indicates no directional influence. That is, when $\Sigma_2 = \Sigma_1$, $F_{Y\to X} = 0$ and Y does not causally affect X, and when $\Sigma_2 < \Sigma_1$, $F_{Y\to X} > 0$, implying there is a causal influence from Y to X. Similarly, the directional GC from X to Y is defined as

$F_{X\to Y} = \ln \frac{\Gamma_1}{\Gamma_2}$. One can also define the instantaneous GC as $F_{X\cdot Y} = \ln \frac{\Gamma_2 \Sigma_2}{\Gamma_2 \Sigma_2 - \Upsilon_2^2}$ to quantify the instantaneous mutual interaction between X and Y. The total GC between X and Y can thereby be defined as $F_{X,Y} = F_{Y\to X} + F_{X\to Y} + F_{X\cdot Y}$. As network reconstruction often deals with a directed graph, one can focus on directional GC $F_{Y\to X}$ and $F_{X\to Y}$ as they quantify the directional causal influence through edges.

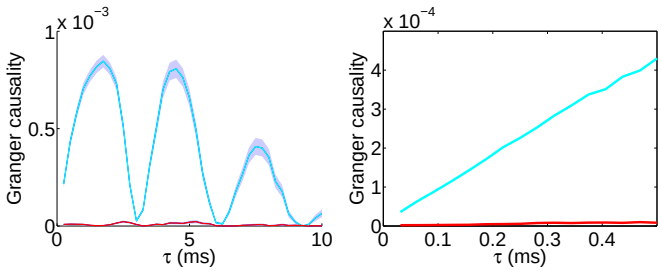
C. Sampling artifacts

As discussed previously, under the setting of uniform sampling, the GC sampling structure (GC value as a function of sampling interval length τ) oscillates and tends to 0 as τ approaches 0 [20, 21]. These phenomena are illustrated in Fig. 1 for time series generated from I&F networks. They affect the reliability of GC inference in the following ways: (i) Directional interactions may be inferred to be absent in certain values of τ; (ii) Spurious GC may arise in certain ranges of τ's; (iii) Finer and finer sampling always gives rise to smaller and smaller GC values. However, for sufficiently fine uniform sampling, one can use a normalized GC (with a scale factor $1/\tau$) to remove the sampling artifacts and obtain a reliable causal inference [20, 21]. In applications, if a sufficiently fine sampling cannot be obtained, the reliability of GC analysis becomes questionable. Therefore, we address the question of how a reliable GC analysis can still be reached if one cannot obtain sufficiently finely sampled time series.

For uniform sampling, if one samples the signal with bandlimit F_0 at rate $F_s < 2F_0$ with equal sampling interval, the aliasing phenomenon can occur, *i.e.,* the spectral component of frequency higher than $\frac{1}{2}F_s$ in the original signal folded into the low frequency ($\leqslant \frac{1}{2}F_s$) component and cannot be distinguished. Therefore, with insufficiently high sampling rate $F_s < 2F_0$, the information in the time series cannot be faithfully extracted and the GC inference becomes unreliable. In contrast, nonuniform sampling can ameliorate the problems caused by aliasing. The reason underlying the aliasing suppression is that for nonuniform sampling there are almost no time domain Fourier bases of different frequencies that can exactly overlap one another on the nonuniform grids [50]. Therefore, we will take advantage of nonuniform sampling schemes to achieve a reliable GC analysis at a lower mean sampling rate than the Nyquist rate. We now turn to the discussion of our approach in the following.

III. NONPARAMETRIC FRAMEWORK FOR NONUNIFORM SAMPLING GC ANALYSIS

The time domain GC, as described in Section II B, is based on linear-regression. In general, it is difficult to generalize this regression procedure to nonuniformly sampled time series. Instead we will use a nonparametric approach to establish a nonuniform sampling GC analysis based on a frequency-domain spectral analysis.



(a) (b)

FIG. 1. (Color online) The GC sampling structure (GC value as a function of sampling interval length τ). The GC analysis is applied to the time series obtained from a two-neuron I&F network with a unidirectional connection from neuron y to neuron x. Plotted are $F_{x\to y}$ (red) and $F_{y\to x}$(cyan) estimated from voltage time series of the system (1) through linear-regression. Here, $\mu = 1\,\mathrm{kHz}$, $\lambda = 0.03$, with coupling strength $s = 0.02$ ($s_{ij} = s$ for $s_{ij} \neq 0$). (a) The oscillatory behavior of GC vs. τ. Shaded regions represent a 95% confidence interval. (b) The vanishing of GC as τ approaches 0. Note that the estimation biases of GC are removed in (b) (See Refs. [20, 21] for details).

A. Notations

First, we introduce basic notions with associated notations to facilitate our discussion. Suppose we have M independent realizations of time-continuous stochastic processes $X^m(t)$, $Y^m(t)$. The power spectral density (PSD) matrix $\mathbf{P}^m(f)$ of the mth realization is defined as

$$\mathbf{P}^m(f) = \begin{bmatrix} TI_x^m I_x^{m*} & TI_x^m I_y^{m*} \\ TI_y^m I_x^{m*} & TI_y^m I_y^{m*} \end{bmatrix}, \tag{3}$$

where T is the length of observation time window, "$*$" denotes complex conjugate transpose. $I_x^m(f) = \frac{1}{T}\int_0^T X^m(t)e^{-\mathrm{i}2\pi ft}\mathrm{d}t$, $I_y^m(f) = \frac{1}{T}\int_0^T Y^m(t)e^{-\mathrm{i}2\pi ft}\mathrm{d}t$, with $f = k/T$, k being an integer, are the Fourier transforms of $X^m(t)$ and $Y^m(t)$, respectively. The average PSD matrix $\widetilde{\mathbf{P}}(f)$ is evaluated from all the realizations as

$$\widetilde{\mathbf{P}}(f) = \frac{1}{M}\sum_{m=1}^{M} \mathbf{P}^m(f).$$

As M tends to infinity, we obtain the empirical PSD matrix $\mathbf{P}(f) = \lim_{M\to+\infty} \frac{1}{M}\sum_{m=1}^{M} \mathbf{P}^m(f)$. Note that $\mathbf{P}(f)$ is still only an approximation to the true PSD. In fact, the empirical $\mathbf{P}(f)$ depends on T and tends to the true PSD matrix as T tends to infinity. Clearly, when T is much larger than the correlation times of X and Y, the difference between the empirical and true PSDs becomes

small. As we choose sufficiently large T, we will no longer distinguish the empirical PSD and the true PSD in the following discussion.

B. Unbiased PSD estimation for nonuniform sampling

For the mth realization of nonuniformly sampled time series $X^m_{t_n}$, $Y^m_{t_n}$, $t_n \sim U(0,T)$, a uniformly distributed random variable, their *nonuniform Fourier transforms* are defined as $I^m_{nu,x}(f) = \frac{1}{N}\sum_{s=1}^{N} X^m_{t_s} e^{-\mathrm{i}2\pi f t_s}$, $I^m_{nu,y}(f) = \frac{1}{N}\sum_{s=1}^{N} Y^m_{t_s} e^{-\mathrm{i}2\pi f t_s}$, where N is the total number of sampling points. The nonuniform mean sampling interval length τ is defined as T/N and the associated rate is N/T. Numerically, the nonuniform Fourier transform can be computed using the nonuniform fast Fourier transform (NUFFT) with the same computational complexity $O(N\log N)$ as the fast Fourier transform (FFT) for uniform grids [51]. Note that $I^m_{nu,x}(f)$ and $I^m_{nu,y}(f)$ converge to $I^m_x(f)$ and $I^m_y(f)$, respectively, as the number of nonuniform sampling points N tends to infinity. For a finite N, $I^m_{nu,x}(f)$, for example, can be regarded as a Monte Carlo approximation to $I^m_x(f)$. Thus, $I^m_{nu,x}(f)$ can be expressed as

$$I^m_{nu,x}(f) = I^m_x(f) + \epsilon^m_{nu,x}(f),$$

where $\epsilon^m_{nu,x}(f) \sim O\left(1/\sqrt{N}\right)$ is the random error function with zero mean at each frequency. Note that $I^m_{nu,x}(f)$, $I^m_x(f)$ and $\varepsilon^m_{nu,x}(f)$ in general are complex. Given the signal of the mth realization $X^m(t)$, $I^m_x(f)$ is fixed on each frequency, $\epsilon^m_{nu,x}(f)$ is determined by the positions of N random sampling points and is independent from the original signal. Estimating the PSD matrix $\mathbf{P}^m(f)$ through a direct replacement of $I^m_x(f)$, $I^m_y(f)$ in Eq. (3) with $I^m_{nu,x}(f)$, and $I^m_{nu,y}(f)$ gives rise to a bias in estimating $\mathbf{P}(f)$. The origin of this bias can be seen as follows. For the power spectrum of X obtained through the direct replacement, $P^m_{DR,xx} = TI^m_{nu,x}I^{m*}_{nu,x}$. Notice that

$$\begin{aligned} P^m_{DR,xx} = {} & TI^m_x I^{m*}_x + T\epsilon^m_{nu,x}I^{m*}_{nu,x} + TI^m_{nu,x}\epsilon^{m*}_{nu,x} \\ & + T\epsilon^m_{nu,x}\epsilon^{m*}_{nu,x}, \end{aligned}$$

and we have

$$\lim_{M\to+\infty}\frac{1}{M}\sum\nolimits_{m=1}^{M} P^m_{DR,xx} = P_{xx} + T\mathrm{var}\left(\varepsilon^m_{nu,x}\right)$$

which is greater than P_{xx}. Here the bias term $T\mathrm{var}\left(\varepsilon^m_{nu,x}\right) = \lim_{M\to+\infty}\frac{T}{M}\sum_{m=1}^{M}\varepsilon^m_{nu,x}\varepsilon^{m*}_{nu,x}$ is proportional to T/N. It can be suppressed by increasing the nonuniform mean sampling rate N/T, but cannot be reduced by increasing the number of realizations M. This fact indicates that, if the bias is not properly removed, even when a large number of realizations are obtained, the PSD estimation is still incorrect and one cannot obtain a reliable GC analysis. Therefore, we propose the following unbiased PSD estimator $\mathbf{P}^m_{nu}(f) = \begin{bmatrix} P^m_{nu,xx}(f) & P^m_{nu,xy}(f) \\ P^m_{nu,yx}(f) & P^m_{nu,yy}(f) \end{bmatrix}$ for the mth realization , where

$$P^m_{nu,ij}(f) = \frac{NT}{N-1}\left(I^m_{nu,i}I^{m*}_{nu,j} - \frac{1}{N^2}\sum_{s=1}^{N} x^m_i(t_s)x^{m*}_j(t_s)\right). \tag{4}$$

Here $i,\ j = x$ or y, $x^m_i(t_s) = X^m_{t_s}$ for $i = x$ and $x^m_i(t_s) = Y^m_{t_s}$ for $i = y$. In Eq. (4), the estimation bias for a PSD function has been subtracted. Note that the bias term $\varepsilon^m_{nu,i}\varepsilon^{m*}_{nu,j}$only appears when i equals j in the above expression of $P^m_{DR,ij}$ and has been subtracted by the difference between $I^m_{nu,i}I^{m*}_{nu,j}$ and $\frac{1}{N^2}\sum_{s=1}^{N} x^m_i(t_s)\, x^m_j(t_s)$ for $i = j$ in Eq. (4) (see Appendix A for details). Therefore, the bias for PSD estimation is removed and we obtain $\mathbb{E}\left(\mathbf{P}^m_{nu}(f)\right) = \mathbf{P}^m(f)$, where $\mathbb{E}$ indicates expectation with respect to different realizations of the N random nonuniform sampling time points for the fixed original signal (see Appendix A for details). After the realization average, the PSD estimator becomes

$$\widetilde{\mathbf{P}}_{nu}(f) = \frac{1}{M}\sum_{m=1}^{M}\mathbf{P}^m_{nu}(f),$$

where, as $M\to+\infty$, $\widetilde{\mathbf{P}}_{nu}(f)\to\mathbf{P}(f)$, which is an unbiased PSD estimator.

C. Spectrum-based nonparametric GC analysis

We now recapitulate the spectrum-based nonparametric GC analysis approach [2, 42]. The central process in this approach is to compute the linear-regression residual covariances from spectrum. The spectral density matrix $\mathbf{S}(\omega)$, $\omega\in[-\pi,\pi]$ can be factorized as $\mathbf{S}(\omega) = \mathbf{A}(e^{-\mathrm{i}\omega})\mathbf{A}^*(e^{-\mathrm{i}\omega})$ [52]. This factorization is unique if (i) $\mathbf{A}(z)$ as a polynomial has only nonnegative powers; (ii) $\det(\mathbf{A}(z))$ and $1/\det(\mathbf{A}(z))$ are analytic inside the unit disk; and (iii) $\mathbf{A}(0)$ is real, upper triangular with positive diagonal elements [52]. The numerical scheme of factorization is proposed by Wilson [52]. It is based on Newton's method. The nth iteration step reads

$$\mathbf{A}_{n+1}(\omega) = \mathbf{A}_n(\omega)\left\{\left[\mathbf{A}_n^{-1}(\omega)\mathbf{S}(\omega)\mathbf{A}_n^{-1*}(\omega) + I\right]^+ + \mathbf{O}_n\right\}, \tag{5}$$

where $\mathbf{S}(\omega)$ is the spectral density matrix with $\omega\in[-\pi,\pi]$, I is the identity matrix. $[\cdot]^+$ is defined as follows. For any matrix function $\mathbf{F}(\omega)$ with $\omega\in[-\pi,\pi]$, suppose it is expanded in frequency domain as $\mathbf{F}(\omega) = \sum_{k=-\infty}^{+\infty}\mathbf{f}_k\mathrm{e}^{\mathrm{i}k\omega}$. Then $[\mathbf{F}(\omega)]^+ = \frac{1}{2}\mathbf{f}_0 + \sum_{k=1}^{+\infty}\mathbf{f}_k\mathrm{e}^{\mathrm{i}k\omega}$. $\mathbf{O}_n$ is a scalar matrix satisfying $\mathbf{O}_n + \mathbf{O}_n^* = 0$ and is uniquely determined as we require $\left[\mathbf{A}_n^{-1}(\omega)\mathbf{S}(\omega)\mathbf{A}_n^{-1*}(\omega) + I\right]^+ + \mathbf{O}_n$ being upper triangle. Usually we can set the initial value

$\mathbf{A}_0(\omega) = I$ for convenience. Note that the convergence order of this factorization method is quadratic in nature [52]. Therefore, $\mathbf{A}_n(\omega)$ will converge to $\mathbf{A}(e^{-\mathrm{i}\omega})$ often in a few steps within a reasonable error tolerance. Finally, we obtain the covariance matrix of the joint-regression residuals through $\mathbf{\Sigma} = \mathbf{A}(0)\mathbf{A}^*(0)$. If $\mathbf{S}$ is a one-by-one matrix, $\mathbf{\Sigma}$ obtained from the above factorization is a scalar, which is the variance of the auto-regression residuals of the corresponding time series. Note that, for the one-by-one matrix case, alternatively, $\mathbf{\Sigma}$ can be computed explicitly by $\mathbf{\Sigma} = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln \mathbf{S}(\omega)\mathrm{d}\omega\right)$.

In principle, high frequency information could be recovered from the nonuniform sampling. However, in practice, one has to truncate $\mathbf{P}(f)$ at a certain frequency f_{cut}. Note that the coherence between time series $X(t)$ and $Y(t)$ is defined as $C(f) = \frac{P_{xy}(f)P_{yx}(f)}{P_{xx}(f)P_{yy}(f)}$ (see Fig. 2c), which is related to the normalized total GC by $\hat{F}_{X,Y} = -\int_{-\infty}^{+\infty} \ln\left[1 - C(f)\right] \mathrm{d}f$ [20]. We can choose a cutoff frequency f_{cut} above which $C(f)$ nearly vanishes. Defining $\omega = \pi f / f_{\mathrm{cut}}$, we can then apply the Wilson's iterative algorithm to factorize the PSD matrix through $\mathbf{S}(\omega) = 2f_{\mathrm{cut}}\mathbf{P}(\omega f_{\mathrm{cut}}/\pi)$. We are now ready to compute GC from $\mathbf{P}(f) = \begin{bmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{bmatrix}$ as follows: (1) Compute the variances Σ_1,Γ_1 of the auto-regression residuals from $P_{xx}(f)$ and $P_{yy}(f)$ using Wilson's iterative factorization algorithm corresponding to one-by-one matrix as discussed above. (2) Compute the covariance matrix $\mathbf{\Sigma} = \begin{bmatrix} \Sigma_2 & \Upsilon_2 \\ \Upsilon_2 & \Gamma_2 \end{bmatrix}$ of the joint-regression residuals from $\mathbf{P}(f)$ using Wilson's iterative factorization algorithm as discussed in the above paragraph. (3) Compute the directional GC as defined in Section II B, namely, $F_{Y\to X} = \ln\frac{\Sigma_1}{\Sigma_2}$, $F_{X\to Y} = \ln\frac{\Gamma_1}{\Gamma_2}$. Note that, because any covariance component obtained through regression can also be obtained using the factorization of the corresponding PSD as illustrated above, such nonparametric GC estimation approach can be naturally extended to conditional GC analysis [10, 53] of multiple neurons as discussed below.

D. GC normalization

For nonuniform sampling GC analysis, the value of the GC computed in Section III C depends on f_{cut}. However, even when f_{cut} is sufficiently large, *i.e.*, above which the coherence nearly vanishes (*e.g.*, $f_{\mathrm{cut}} \geqslant 1\,\mathrm{kHz}$ in Fig. 2c), GC decreases slowly as f_{cut} tends to infinity as shown in Fig. 2a. Theoretically, the slow decay of nonuniform sampling GC as a function of f_{cut} is related to the scaling paradox in which uniform sampling GC vanishes linearly as the sampling interval length τ approaches 0. Here, for nonuniform sampling, we can define an effective uniform sampling interval length as $\frac{1}{2f_{\mathrm{cut}}}$. As is illustrated in the inset of Fig. 2a, GC vanishes linearly as $\frac{1}{2f_{\mathrm{cut}}}$ tends to 0. In Refs. [20, 21], GC is normalized by $1/\tau$ to obtain a convergent normalized GC value as τ tends to 0. This normalized GC approaches a nonzero constant as τ tends to 0 if there is a causal influence or zero, otherwise. Here, similarly, we need to normalize GC as follows,

$$\begin{aligned} \hat{F}_{Y\to X} &= 2f_{\mathrm{cut}} \ln\frac{\Sigma_1}{\Sigma_2}, \\ \hat{F}_{X\to Y} &= 2f_{\mathrm{cut}} \ln\frac{\Gamma_1}{\Gamma_2}, \end{aligned} \tag{6}$$

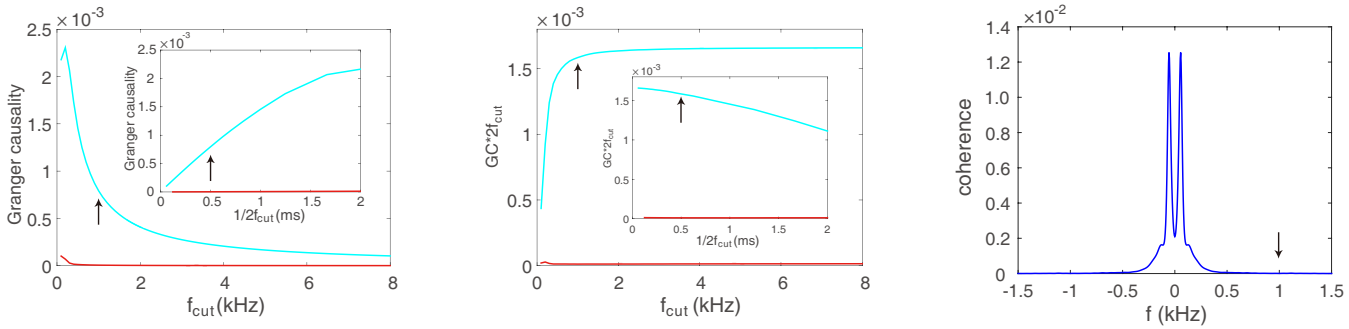
which become nearly invariant to the cutoff frequency f_{cut} when f_{cut} becomes sufficiently large, *e.g.*, $f_{\mathrm{cut}} \geqslant 1\,\mathrm{kHz}$ ($\frac{1}{2f_{\mathrm{cut}}} \leqslant 0.5\,\mathrm{ms}$) as shown in Fig. 2b.

If there is a causal influence, this normalized GC approaches a nonzero constant, otherwise, it approaches zero, as f_{cut} tends to infinity. The normalization procedure applies to any GC analysis (uniform or nonuniform) on time series sampled from the general type of time-continuous quantities because such a scaling property is intrinsic for GC analysis [20, 21]. Therefore, for uniformly or nonuniformly sampled time series, no matter whether it is parametric or nonparametric, we should always perform the normalization $\hat{F} = 2f_{\mathrm{cut}}F$ (For the case of uniform sampling, $2f_{\mathrm{cut}}$ is usually set to $\frac{1}{\tau}$, where τ is the uniform sampling interval length).

To summarize, we have described a spectrum-based nonparametric framework of GC inference for nonuniformly sampled time series. This framework is general and can be directly extended to any sampling scheme and any type of time series by applying a proper method of spectrum estimation.

IV. SPECTRAL PROCESSING

The nonparametric GC framework presented in Section III extends the GC analysis to the case of nonuniformly sampled data. However, a direct application of this approach may result in a significant estimation error. This error arises from (i) the statistical error of the nonuniform sampling PSD, which remains at a nearly constant level at high frequencies where the true PSD almost vanishes (See Fig. 3), and (ii) the spectrum-based factorization procedure, which can greatly magnify the error of the GC estimated from the PSD. In addition, when the error is greater than the magnitude of the auto PSD, the unbiased spectral estimation approach sometimes can give rise to negative values of auto PSD at some frequencies as opposed to the fact that an auto PSD should be always nonnegative. Then the factorization procedure may fail because it requires a positive definite spectrum. Therefore, if the data size is not sufficiently large to suppress statistical errors in estimating PSD, one may not achieve good performance of the nonparametric GC framework. In order to reduce the requirement of data size and at the same time improve the estimation accuracy, we attempt to devise a specific spectral processing by incorporating characteristics of the PSD of



(a) (b) (c)

FIG. 2. (Color online) The dependence of GC on the cutoff frequency $f_{\rm cut}$ for nonuniform sampling. (a) GC v.s. $f_{\rm cut}$ and GC v.s. $1/2f_{\rm cut}$ (inset), (b) normalized GC v.s. $f_{\rm cut}$ and normalized GC v.s. $1/2f_{\rm cut}$ (inset). Plotted are (a) $F_{y\to x}$ (cyan), $F_{x\to y}$ (red), and (b) $2f_{\rm cut}F_{y\to x}$ (cyan), $2f_{\rm cut}F_{x\to y}$ (red) as a function of $f_{\rm cut}$. (c) Coherence as a function of f. For Figs. 2-6, the time series are generated from the same two-neuron I&F network with a unidirectional connection from neuron y to neuron x with parameters $\mu = 1\,$kHz, $\lambda = 0.012$, and coupling strength $s = 0.02$ ($s_{ij} = s$ for $s_{ij} \neq 0$). Note that here we use PSD functions that are accurately evaluated from a sufficiently long time series (10^8 ms) and sufficiently small sampling interval length (0.0625 ms) for GC analysis. These PSD functions are regarded as the empirical PSD. The black arrows in (a)$\sim$(c) point to $f_{\rm cut} = 1\,$kHz.

the analyzed data. As an example, we use the nonlinear pulse-coupled I&F network dynamics (1) to illustrate our approach. We note that the spectrum of I&F dynamics P_{yy} as shown in Fig. 3a resembles a Lorentzian with a bump at certain frequency, which can be observed in a variety of neural network stochastic processes, e.g., linearized Wilson-Cowan-type equations [54, 55]. Therefore, our analysis for PSD of I&F dynamics is rather general and may potentially be applied to other types of neural dynamics. In the following, we discuss in detail the features of the PSD functions of the I&F networks and take advantage of such features in our spectral processing for error reduction.

A. Decay rate and discontinuity

By the Wiener-Khintchin theorem, the spectral density function of a time series is the Fourier transform of its covariance function. For the PSD functions of the I&F networks, we observe that they are smooth functions of frequency and exhibit power-law decay at high frequencies (see Fig. 3). These features in the Fourier transform result from the continuity and long time behaviors in the covariance function. Before discussing this in detail, we will briefly state the following facts. First, we define the function $H_n(x)$, $n = 1, 2, \cdots$,

$$H_1(x) = \begin{cases} 1/2 & \text{for } x \geqslant 0 \\ -1/2 & \text{for } x < 0 \end{cases},$$

$$H_{n+1}(x) = \int_0^x H_n(s)\mathrm{d}s = \begin{cases} \frac{1}{2n!}x^n & \text{for } x \geqslant 0 \\ -\frac{1}{2n!}x^n & \text{for } x < 0 \end{cases},$$

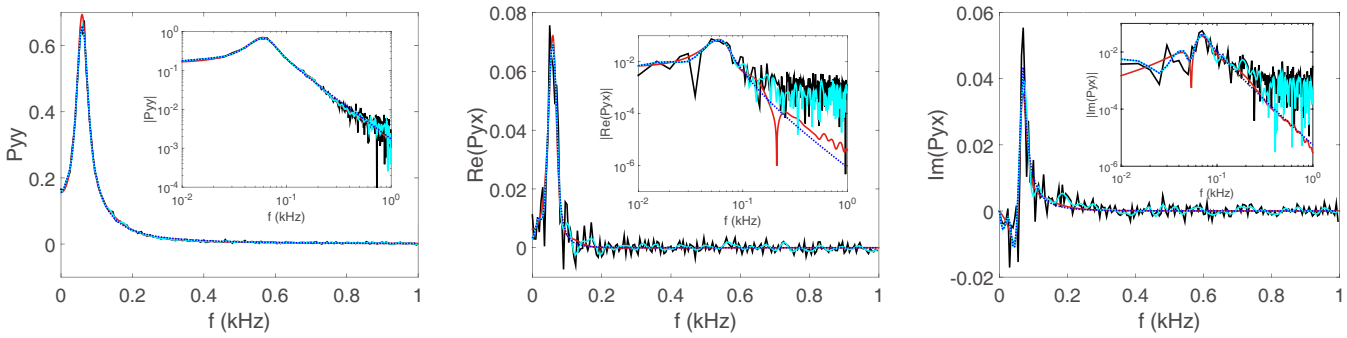
for $n = 1, 2, \cdots$. $H_n(x)$ is a smooth function of x except for the discontinuity of the $(n-1)$th derivative at the origin. The corresponding Fourier transform [56] of $H_n(x)$ for $f \neq 0$ is

$$\mathcal{F}\left(H_1(x-a)\right) = \int_{-\infty}^{+\infty} H_1(x-a)e^{-\mathrm{i}2\pi fx}\mathrm{d}x = \frac{1}{\mathrm{i}2\pi f}\mathrm{e}^{-\mathrm{i}2\pi af},$$

$$\mathcal{F}\left(H_n(x-a)\right) = \frac{1}{(\mathrm{i}2\pi f)^n}\mathrm{e}^{-\mathrm{i}2\pi af}. \tag{7}$$

From Eq. (7), it can be easily seen that the discontinuity of the $(n-1)$th derivative in a function of x leads to a power law decay f^{-n} in its Fourier transform. In addition, for a function with its discontinuity occurring at the origin, *i.e.*, $a = 0$, in Eq. (7), its Fourier transform decays without oscillations at the high-frequency tail. As is well known, if a function is smooth, *i.e.*, all orders of its derivatives are continuous, its Fourier transform decays more rapidly than any power-law function as the frequency f tends to infinity. Conversely, if a function decays more rapidly than any power-law function, *e.g.*, an exponential decay, its Fourier transform (or inverse Fourier transform) is smooth.

According to the above relations, instead of directly investigating the PSD behaviors of the I&F network model, we can focus our study on the corresponding covariance matrix, which for the bivariate time series, X_t and Y_t, is expressed as $\mathbf{C}(\tau) = \begin{bmatrix} C_{xx}(\tau) & C_{xy}(\tau) \\ C_{yx}(\tau) & C_{yy}(\tau) \end{bmatrix}$, where $C_{xx}(\tau)$ and $C_{yy}(\tau)$ are the auto-covariances of time series, X_t and Y_t, respectively, $C_{xy}(\tau)$ and $C_{yx}(\tau)$ are their cross-covariances. The covariance is defined as



(a) (b) (c)

FIG. 3. (Color online) Comparison between estimated PSDs: empirical PSDs (red), nonuniform sampling PSDs (black), nonuniform sampling PSDs after covariance truncation (cyan), nonuniform sampling PSDs after covariance truncation and high-frequency power-law tail-fitting (dotted blue). (a) The auto PSD for neuron y: P_{yy}. (b) The real part of the cross PSD: $\mathrm{Re}\,(P_{yx})$. (c) The imaginary part of the cross PSD: $\mathrm{Im}\,(P_{yx})$. Insets are the corresponding log-log plots of their absolute values. The PSDs are estimated from X_t and Y_t generated by a two-neuron I&F network with a unidirectional connection from neuron y to neuron x.

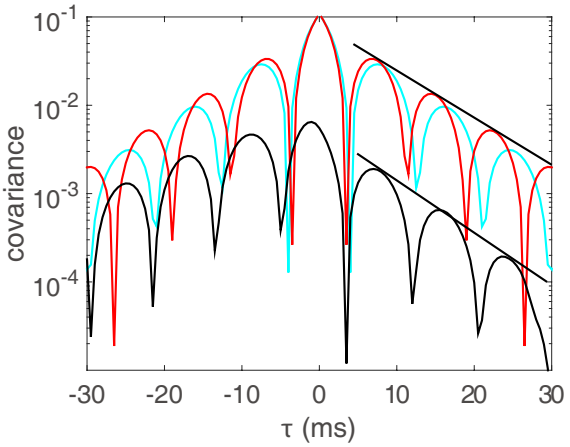


FIG. 4. (Color online) Exponential decay of covariances. $|C_{xx}|$ (red), $|C_{xy}|$ (black), $|C_{yy}|$ (cyan) are estimated from bivariate time series X_t, Y_t generated by a two-neuron I&F model with a unidirectional connection from neuron y to neuron x.

$C_{ij}(\tau) = \mathbb{E}_t\left(x_i(t)x_j(t-\tau)\right) - \mathbb{E}_t\left(x_i(t)\right)\mathbb{E}_t\left(x_j(t)\right)$ for any time series x_i and x_j where $\mathbb{E}_t$ is the expectation with respect to time. Note that the PSD matrix $\mathbf{P}(f)$ is related to the covariance matrix $\mathbf{C}(\tau)$ through the Fourier transform as

$$\mathbf{P}(f) = \int_{-\infty}^{+\infty} \mathbf{C}(s)e^{-\mathrm{i}s2\pi f}\mathrm{d}s.$$

B. Decay of the covariance functions and their truncation

As shown in Fig. 3, which plots the PSD functions of time series generated by an I&F network, the empirical PSDs are smooth whereas the nonuniform sampling PSDs fluctuate strongly. As will be discussed below, these fluctuations could contribute considerably to the GC estimation errors. Therefore, if we could reduce the fluctuations of the nonuniform sampling PSDs properly, we may be able to obtain a better estimation of PSDs, thus giving rise to a reliable spectrum-based GC inference.

Following previous discussions, we examine the decay behavior of the high-frequency tail of the corresponding covariance function. In Fig. 4, an oscillatory exponential decay behavior is observed for all the covariance functions indicating that all the corresponding PSDs are smooth. Incidentally, this exponential decay behavior in covariance functions implies that the I&F network (1) has a finite memory. For the nonuniform sampling scheme, the PSDs possess highly fluctuating errors. These errors will reflect in the erroneous long time behavior of the corresponding covariances whereas the true covariances should approach 0 at large times. This fact naturally suggests a covariance truncation scheme to obtain smooth PSDs, thus leading to reliable GC inference. This can be carried out as follows:

Step 1: Obtain $\mathbf{C}(\tau)$ by applying the inverse Fourier transform to $\mathbf{P}(f)$, that is, $\mathbf{C}(\tau) = \int_{-\infty}^{+\infty} \mathbf{P}(f)e^{\mathrm{i}\tau 2\pi f}\mathrm{d}f$.

Step 2: Choose a proper cutoff time τ_{cut} according to the dynamics so that $\mathbf{C}(\tau)$ nearly vanishes for $|\tau| > \tau_{\mathrm{cut}}$. For any specific data, we can also choose τ_{cut} such that, for $|\tau| > \tau_{\mathrm{cut}}$, the estimated covariance is essentially dominated by statistical fluctuations.

Step 3: Define the truncated covariance matrix $\tilde{\mathbf{C}}(\tau) =$

$\mathbf{C}(\tau)$ for $|\tau| \leqslant \tau_{\rm cut}$ and $\tilde{\mathbf{C}}(\tau) = 0$ otherwise.

Step 4: Obtain a smoothed PSD $\mathbf{P}_{\rm s}(f)$ by applying the Fourier transform to the truncated covariance matrix $\tilde{\mathbf{C}}(\tau)$, *i.e.*, $\mathbf{P}_{\rm s}(f) = \int_{-\infty}^{+\infty} \tilde{\mathbf{C}}(s) e^{-{\rm i}s2\pi f} {\rm d}s$.

Note that the truncation of the covariance may lead to a jump discontinuity, thus affecting the decay behavior of the corresponding PSD. However, as long as the jump of the discontinuity is much smaller than the magnitude of the covariance estimation noise, the effect of this discontinuity on the corresponding PSD estimation can be neglected. In addition, we note that the exponential decay behavior in covariance is not specific to the I&F network dynamics, it is rather general for some other dynamics as well. For example, for the widely used auto-regressive model in time series analysis, the linear iterative process naturally generates a time series with covariance possessing an exponential decay behavior. Therefore, our truncation scheme can also be applied to a broad class of dynamics to reduce the PSD estimation error.

C. Covariance discontinuity and power-law tail-fitting

As discussed in Section IV A, to analyze the decay behavior of the PSD of the I&F network dynamics, we can first study the discontinuity in the corresponding covariance function. By invoking the above discussed features of the covariance function, we devise a high-frequency power-law tail-fitting scheme for error reduction in estimating the PSD.

1. *Discontinuity of the auto-covariance function*

For the auto-covariance $C_{xx}(\tau)$ of the I&F dynamics, as shown in Fig. 5, a discontinuity of its first derivative can be clearly observed at the origin. As will be demonstrated below, such a discontinuity in $\frac{{\rm d}C_{xx}(\tau)}{{\rm d}\tau}$ is generic for the I&F dynamics. We can compute the derivative of $C_{xx}(\tau)$ at the origin, $\left.\frac{{\rm d}C_{xx}(\tau)}{{\rm d}\tau}\right|_{0^+} = \lim_{\tau\to 0^+} \frac{\rm d}{{\rm d}\tau}\mathbb{E}_t\left[\left(X(t)-m_x\right)\left(X(t-\tau)-m_x\right)\right]$, where $X(t)$ is the dimensionless voltage trace, $m_x = \mathbb{E}_t\left(X(t)\right)$ is the mean value of $X(t)$, $\mathbb{E}_t$ is the expectation with respect to time. Note that the voltage time series can be regarded as being wide-sense stationary, therefore, the ensemble average is equivalent to the time average for the second order statistics, *i.e.*, covariance. In the following, we do not distinguish between these two types of averages. After a careful treatment of the jump discontinuity in $X(t)$ (see Appendix B for details), we obtain

$$\left.\frac{{\rm d}C_{xx}(\tau)}{{\rm d}\tau}\right|_{0^+} = -\frac{1}{2}\left(x^{\rm th}-x^{\rm r}\right)^2 R_x, \tag{8}$$

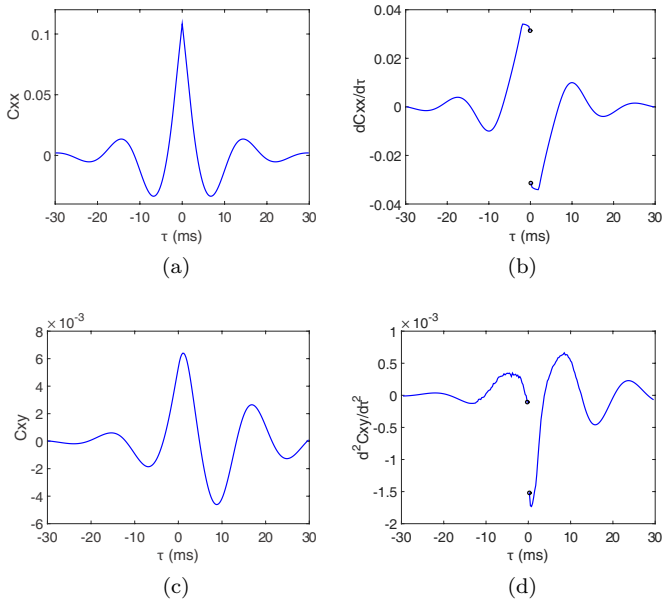


FIG. 5. (Color online) Covariances and their derivatives. (a) Auto-covariance $C_{xx}(\tau)$, (b) $\frac{{\rm d}C_{xx}(\tau)}{{\rm d}\tau}$, (c) cross-covariance $C_{xy}(\tau)$, (d) $\frac{{\rm d}^2C_{xy}(\tau)}{{\rm d}\tau^2}$, computed from the voltage time series $X(t)$ and $Y(t)$ of neuron x and neuron y, respectively, which are generated by a two-neuron I&F network with a unidirectional connection from neuron y to neuron x. From the full simulation of the I&F dynamics of neuron x, we obtain the firing rate $R_x \approx 0.065\,{\rm ms}^{-1}$. From numerical computation of the auto-covariance $C_{xx}(\tau)$, we have $\left.\frac{{\rm d}C_{xx}(\tau)}{{\rm d}\tau}\right|_{0^+} = -\left.\frac{{\rm d}C_{xx}(\tau)}{{\rm d}\tau}\right|_{0^-} = -0.032\,{\rm ms}^{-1}$, which is equal to $-\frac{1}{2}\left(x^{\rm th}-x^{\rm r}\right)^2 R_x$ within our numerical accuracy.

where R_x is the mean firing rate of neuron x. Since $C_{xx}(\tau)$ is an even function of τ, Eq. (8) yields

$$\left.\frac{{\rm d}C_{xx}(\tau)}{{\rm d}\tau}\right|_{0^+} - \left.\frac{{\rm d}C_{xx}(\tau)}{{\rm d}\tau}\right|_{0^-} = -\left(x^{\rm th}-x^{\rm r}\right)^2 R_x, \tag{9}$$

which is the jump size of the first derivative discontinuity of $C_{xx}(\tau)$ at the origin. An example is illustrated in Fig. 5b.

2. *Discontinuity of the cross-covariance function*

An example of the cross-covariance generated by a specific I&F network is shown in Fig. 5c. From Fig. 5d, it can be clearly seen that there is a discontinuity in the second derivative of $C_{xy}(\tau)$ at the origin. Continuing the discussion of our I&F model with a unidirectional connection from neuron y to neuron x with coupling strength s, we discuss how the discontinuity of $\frac{{\rm d}^2C_{xy}(\tau)}{{\rm d}\tau^2}$ arises from the I&F dynamics. By definition, we have $\frac{{\rm d}^2C_{xy}(\tau)}{{\rm d}\tau^2} = \frac{{\rm d}^2}{{\rm d}\tau^2}\mathbb{E}_t\left(X(t)Y(t-\tau)\right)$, where $X(t)$ and $Y(t)$ are dimensionless voltage time series for neuron x and

neuron y, respectively. By translational invariance in time (see Appendix C for details), we can obtain

$$\frac{\mathrm{d}^2 C_{xy}(\tau)}{\mathrm{d}\tau^2} = -\mathbb{E}_t\left(X'(t)Y'(t-\tau)\right), \tag{10}$$

where the prime denotes the derivative with respect to time. For the I&F dynamics, the voltage x_i of neuron i is reset from x^{th} to x^{r} at each firing time $T_{i,j}$, $j = 1, 2, \cdots$. Therefore, for the voltage trace x_i, there are jump discontinuities at $T_{i,j}$ with jump magnitude $-\left(x^{\mathrm{th}} - x^{\mathrm{r}}\right)$, which lead to singularities in x_i' as described by $-\left(x^{\mathrm{th}} - x^{\mathrm{r}}\right)\delta(t - T_{i,j})$. From the definition (2), the spike train time series is described as $S_T^i(t) = \sum_j \delta(t - T_{i,j})$. Therefore, we can separate $X'(t)$ and $Y'(t)$ as $X'(t) = X_c'(t) - \left(x^{\mathrm{th}} - x^{\mathrm{r}}\right) S_T^x(t)$, $Y'(t) = Y_c'(t) - \left(x^{\mathrm{th}} - x^{\mathrm{r}}\right) S_T^y(t)$, where $X_c'(t)$ and $Y_c'(t)$ are the nonsingular part of $X'(t)$ and $Y'(t)$, respectively. From our numerical experiments, the second derivative discontinuity in cross-covariance is mainly contributed by the interplay between the singular parts $S_T^x(t)$ and $S_T^y(t)$. Therefore, we can concentrate on the singular parts of $X'(t)$ and $Y'(t)$, *i.e.*, $-\left(x^{\mathrm{th}} - x^{\mathrm{r}}\right) S_T^x(t)$ and $-\left(x^{\mathrm{th}} - x^{\mathrm{r}}\right) S_T^y(t)$, to understand the discontinuous behavior of $\frac{\mathrm{d}^2 C_{xy}(\tau)}{\mathrm{d}\tau^2}$. Because x^{th} and x^{r} are constants, we only need to compute $\mathbb{E}_t\left(S_T^x(t) S_T^y(t-\tau)\right)$ to describe the jump in the discontinuity. Through mathematical derivation (see Appendix C for details), we can demonstrate that, $\mathbb{E}_t\left(S_T^x(t) S_T^y(t-\tau)\right)\big|_{\tau\to 0^+} - \mathbb{E}_t\left(S_T^x(t) S_T^y(t-\tau)\right)\big|_{\tau\to 0^-} > 0$, hence $\left.\frac{\mathrm{d}^2 C_{xy}(\tau)}{\mathrm{d}\tau^2}\right|_{0^+} - \left.\frac{\mathrm{d}^2 C_{xy}(\tau)}{\mathrm{d}\tau^2}\right|_{0^-} < 0$. Therefore, in a wide range of dynamical regimes where the singular parts of $X'(t)$ and $Y'(t)$ dominate the discontinuity behavior, we can conclude that $\left.\frac{\mathrm{d}^2 C_{xy}(\tau)}{\mathrm{d}\tau^2}\right|_{0^+} - \left.\frac{\mathrm{d}^2 C_{xy}(\tau)}{\mathrm{d}\tau^2}\right|_{0^-}$ does not vanish and there is a jump discontinuity in the second derivative of the cross-covariance at the origin (Fig. 5d). Note that, from the mathematical derivation in Appendix C, the discontinuity in $\frac{\mathrm{d}^2 C_{xy}(\tau)}{\mathrm{d}\tau^2}$ is determined by the coupling structure in the I&F dynamics (1). As expected, if the firing event of neuron y leads to a continuous change in the conductance of neuron x instead of an instantaneous increment as in the dynamics (1), the discontinuity of cross-covariance may occur at a higher order derivative.

3. *High-frequency tail-fitting in PSDs*

Through the analysis of discontinuities in covariance functions for the time series obtained from the I&F dynamics, we can conclude that, for a wide range of dynamical regimes, (i) the auto-covariance function possesses a discontinuity in its first derivative at the origin, (ii) the cross-covariance function possesses a discontinuity in its second derivative at the origin. These discontinuities give rise to the following properties for the PSD functions for the I&F dynamics:

1. The auto PSD functions $P_{xx}(f)$ and $P_{yy}(f)$ have a power law decay as f^{-2} at high frequencies.

2. The cross PSD function $P_{xy}(f)$ has a power-law decay as f^{-3} at high frequencies.

3. The PSD functions do not oscillate at high frequencies because all their discontinuities occur at the origin.

Before discussing how we can take advantage of these decay features to reduce the PSD estimation errors, we first examine the general behavior of PSD before crossing over to the high frequency decay. For a wide range of I&F dynamics, we have observed the following phenomena in our two-neuron I&F network dynamics:

1. An auto PSD function usually has only one broad peak. This peak is located approximately at the mean firing rate of the corresponding neuron.

2. The cross PSD is complex. Both the real and imaginary parts of the cross PSD are smooth. They usually have one main peak located between the mean firing rates of two neurons.

Based on the above observations, we fit the high-frequency tail as follows:

Step 1. For a given PSD function, locate the maximum value of its absolute value, denoted by $|P_{\max}|$, and the corresponding peak frequency f_{P}.

Step 2. Find the largest frequency f_{st} satisfying $|P(f_{\mathrm{st}})| > \alpha\,|P_{\max}|$, where α is a parameter $0 < \alpha < 1$. f_{st} is used to separate the high-frequency tail of the given PSD function from its peak. The parameter α should be chosen judiciously so as to include sufficiently broad high frequency tail of the PSD function while ensuring its tail for $f > f_{\mathrm{st}}$ to maintain a consistent power-law decay behavior. Usually, for the I&F dynamics, the fitting result is not sensitive to the choice of α as long as it is in an appropriate range, e.g., $0.6 < \alpha < 0.8$.

Step 3. Using the least-square (LS) method to fit the PSDs in the interval $[f_{\mathrm{st}}, f_{\mathrm{cut}}]$ with a function that has asymptotically a power-law decay at high frequencies. Note that, as the functions we use for fitting is nonlinear with respect to parameters, an iterative trust-region approach [57] is applied to the LS optimization. For an auto PSD, we can use the rational function $\frac{a}{x^2+bx+c}$, where a, b, and c are parameters determined by the LS method. For a cross PSD, which is a complex function, we can use the complex rational function $\frac{a_1}{(x+c_1)^3} + \frac{a_2}{(x+c_2)^3}\mathrm{i}$ to fit its real and imaginary parts. As will be discussed below, these rational functions determined by the LS fitting can well capture the high-frequency tail behavior of PSD. There is no need to choose a higher order rational function to fit the high-frequency tail behavior of PSD because the PSD estimation errors can be large and the LS fitting could lead to an incorrect result similar to the Runge phenomenon when using higher order rational functions [58, 59].

Here, the power-law tail-fitting approach is not specific to the I&F dynamics (1). In general, it can be applied to dynamics, which has discontinuity in derivatives thus a power-law decay in the tail of its PSD, to reduce statistical fluctuations in the tail of its PSD. For auto PSD of a

dynamical variable, the decay order of power-law in the tail is determined by the discontinuity in its trajectory, whereas for cross PSD between two dynamical variables, it is determined by the discontinuity in both their trajectories and the interactions between them.

Note that the above spectral processing procedures can also be applied to PSD matrix with respect to multiple neurons for conditional GC analysis. In the following section, we will see that, with our procedures, PSD estimation errors in the nonuniform sampling GC analysis can be greatly reduced.

V. NUMERICAL RESULTS

In the following, we use numerical experiments to demonstrate the effectiveness of our nonuniform sampling GC analysis and spectral processing approaches. Note that, to mimic the nonuniform sampling process, we first numerically evolve the I&F network dynamics using very fine time-step interval, e.g., $\delta t = 1/16$ms, to obtain a simulated uniformly spaced time series (transient effects have been removed) representing the underlying time-continuous dynamics X_t where $t = \delta t, 2\delta t, \cdots, T$. Then, the nonuniformly sampled time series X_{t_n} is obtained from X_t by generating each t_n from a discrete uniform distribution on the whole time-step interval set $\{\delta t, 2\delta t, \cdots, T\}$. In addition, we have verified that our numerical results reported throughout are not sensitive to the choice of initial conditions.

A. Effects of the spectral processing

To examine how well our procedure can reduce errors in estimating PSD or covariances when applied to the nonlinear I&F network data, we compare the PSDs through each step of the spectral processing with the empirical PSD as shown in Fig. 3. It can be clearly seen that: (i) the nonuniform sampling PSD fluctuates strongly around the true PSD and such fluctuations do not decay at high frequencies; (ii) the covariance truncation smooths the original nonuniform sampling PSD by reducing the fluctuation amplitude of PSD in both low and high frequency parts, but high-frequency tail part still does not decay; (iii) the high-frequency power-law tail-fitting can well capture the decay behavior of the PSD at high frequencies, thus greatly improving the spectral estimation accuracy at high frequencies. Clearly, our procedures allow one to obtain a rather precise and accurate PSD estimation from the original nonuniform sampling PSD, which is highly fluctuating with non-vanishing statistical errors at high frequencies and is not suitable for a reliable GC analysis.

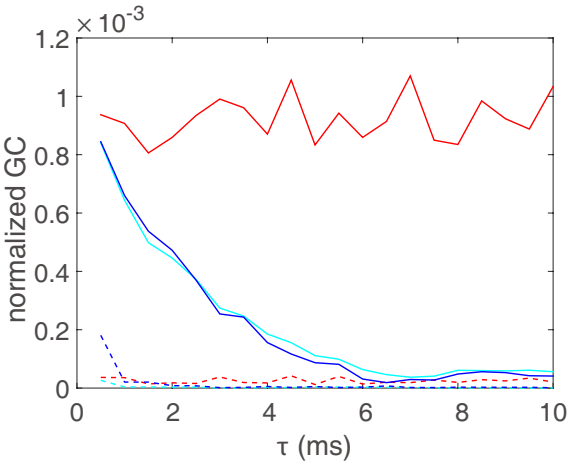


FIG. 6. (Color online) Comparison between the normalized GC sampling structures. $F_{x\to y}$ (dash), $F_{y\to x}$ (solid) are obtained by the conventional linear-regression (cyan), by the uniform sampling nonparametric GC analysis (blue), by the nonuniform sampling GC analysis with truncation and fitting procedures (red). The time series are generated by a two-neuron I&F network with a unidirectional connection from neuron y to neuron x. Note that, for the evaluation of each normalized GC point in the figure, we use exactly the same number of data points of the time series for every method.

B. The GC sampling structures

For uniformly sampled data, as discussed previously, the GC sampling structures (GC value as a function of sampling interval length τ) oscillate as the sampling interval length τ increases. We now turn to the demonstration of the removal of these sampling artifacts by applying the nonuniform sampling scheme together with our PSD processing procedures. Note that the normalized GC (Eq. (6)) is always used in the following discussion.

In Fig. 6, using the same number of data points, we compare the normalized GC sampling structures obtained through different methods. The nonuniform GC sampling structure stays nearly as a constant even for large τ. For sufficiently small τ, both the uniform sampling GC (regression-based or nonparametric) and the nonuniform sampling GC obtained with the truncation and fitting procedures can produce similar values for causal inference. However, for large τ, the uniform sampling GC analysis yields very small normalized GC values at a certain range of τ indicating that there seems no coupling in the network. In other words, it becomes difficult to distinguish the coupled direction from the uncoupled direction at these τs. Note that for uniformly sampled time series the nonparametric GC analysis produces a similar GC sampling structure to the regression-based GC analysis [20, 21]. Therefore, a general nonparametric GC analysis may not always provide a reliable inference for large τ. In contrast, GCs estimated from nonuniform PSDs with the truncation and fitting procedures can well distinguish the coupled direction from the un-

coupled even when the mean sampling interval length τ is large. Clearly, the nonuniform sampling GC analysis obtained through our spectral processing can overcome the sampling artifacts of the uniform sampling GC analysis (no matter parametric or nonparametric) and produce reliable causal inference.

In the case of multiple neurons' time series $\{x_i(t)\}_{i=1}^N$ with $N > 2$, the causal relation between two time series, say, $x_1(t)$ and $x_2(t)$, can be directly mediated or it can be indirectly mediated by a third one, say $x_3(t)$. To determine whether the causal influence is direct, the framework of conditional GC was developed [53], where "conditional" means both the auto-regression and joint regressions of $x_i(t)$ and $x_j(t)$ are performed when the history of all other time series $\{x_k(t)\}$ $(k \neq i, j)$ is given. Similarly, the spectrum-based nonparametric conditional GC analysis was also developed through Wilson's iterative factorization algorithm [42]. We note that our nonuniform GC framework can be naturally extended to the conditional GC analysis for networks of mulple neurons. Fig. 7 displays a numerical example, in which conditional GC analysis combined with truncation and fitting procedures is applied to a ten-neuron I&F network. As discussed in our previous works [17, 18], there is an interesting phenomenon as observed for the GC values of I&F network dynamics (1): if we rank the GC by magnitude for all possible directed connections between neurons, there is a gap separating these ranked GC values as indicated by the black horizontal line in Fig. 7b. This gap clearly divides the GC values into two distinct groups. By using this gap and, for example, choosing a horizontal line within the gap as the GC threshold, the coupled directions can be distinguished from the uncoupled directions and the network topology can be well reconstructed.

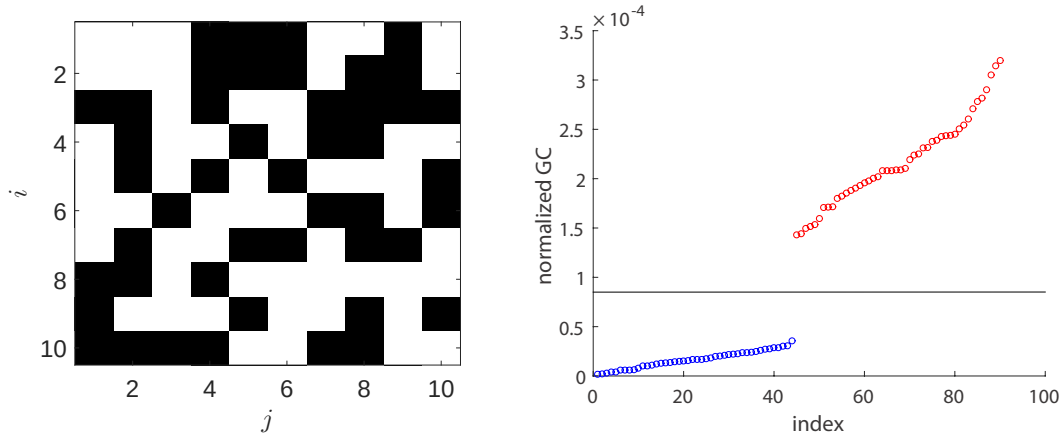
VI. DISCUSSION AND CONCLUSION

In summary, the procedures to perform GC analysis for nonuniformly sampled data is in general as follows: (i) perform nonuniform fast Fourier transform on nonuniformly sampled time series and remove the bias term to obtain an unbiased power spectral density estimation as discussed in Section III B; (ii) use Wilson's iterative factorization algorithm on the estimated power spectral density to obtain the variances of auto-regression and joint-regression residuals as discussed in Section III C; (iii) compute the GC value from the definition as discussed in Section II B.

In this work, we have validated the GC sampling structures in our nonuniform sampling GC framework and have shown that our nonuniform sampling approach can indeed circumvent the sampling artifacts inherent in the uniform sampling GC analysis. Reliable GC inference can be achieved for nonuniformly sampled time series regardless of whether the nonuniform mean sampling rate is high or low. We have also pointed out that for uniform sampling it is the aliasing that leads to the oscillatory sampling artifacts. Furthermore, in applications, our method allows us to nonuniformly sample the signal below the Nyquist rate while achieving a reliable GC analysis. This may potentially overcome the sampling rate limitation on certain devices. In our nonparametric nonuniform sampling GC framework, to suppress the PSD estimation error, we have introduced the truncation and fitting schemes. Our spectral processing is based on the structures of PSD for the I&F network dynamics. As these structures are rather general, our approach can be easily extended to other dynamics to obtain high-quality PSD estimation. Finally, using the same number of data points in the time series for different GC inference methods, in comparison to the uniform sampling GC analysis approaches (regression-based or nonparametric), our nonuniform sampling GC analysis can achieve more reliable and accurate inferences at a lower mean sampling rate than the Nyquist rate.

In addition, we demonstrate that, besides the nonuniform sampling scheme discussed in our work, a nonuniformly sampled time series could be generated in many other different ways. For example, one could take a sufficiently finely sampled time series from a dynamical process with every time point selected with some probability p, rejected with probability $1 - p$, or the nonuniformly sampled time series could be generated by the data collection method of certain experimental device. Our work here is concentrated on how to develop a general framework of spectrum-based reliable GC analysis for given nonuniformly sampled time series. The central point of the framework is to obtain good estimation of the power spectral density of nonuniformly sampled data. As one example of applications of our framework, we investigate the integrate-and-fire (I&F) network dynamics and illustrate how to utilize the particular spectral structure possessed by I&F network dynamics to obtain good estimation of the power spectral density, thus giving rise to reliable GC inference. In real applications, there may be certain constraints for data measurement, e.g., data length or maximum data resolution, and it will be interesting to investigate how to obtain good estimation of the power spectral density for a specific nonuniform sampling method and compare the GC results with different nonuniform sampling methods under these constraints.

Finally, we point out that there is another nonparametric method called transfer entropy which has been formulated to detect directed information transfer in dynamical processes. It has since rapidly gained popularity, particularly in neuroscience, as a tool for data-driven detection of functional connectivity between dynamical variables [60]. For linear Gaussian processes, it has been shown that the transfer entropy and the GC are equivalent up to a factor of 2 [61]. We have demonstrated in our previous works that the uniform sampling artifacts are not specific to the nonlinear I&F dynamics and they also occur in linear Gaussian processes [20, 21]. Therefore, the transfer entropy will also suffer from the uniform sampling artifacts, at least for linear Gaussian processes.



(a) (b)

FIG. 7. (Color online) Conditional nonuniform GC analysis for a ten-neuron I&F network. The parameters are $\mu = 1\,\text{kHz}$, $\lambda = 0.01$, with coupling strength $s = 0.01$ ($s_{ij} = s$ for $s_{ij} \neq 0$). (a) The adjacency matrix A_{ij} of the network. $A_{ij} = 0$ (white) when $s_{ij} = 0$, and $A_{ij} = 1$ (black) when $s_{ij} \neq 0$. (b) Ranked normalized GC values obtained through conditional nonuniform GC framework with truncation and fitting procedures. The nonuniform mean sampling interval length is $\tau = 10\,\text{ms}$. Blue circles indicate the GC values for uncoupled directions whereas red circles indicate the GC values for coupled directions. The solid black line indicates a natural GC threshold (8.5×10^{-5}), below which the corresponding direction is inferred as uncoupled. By applying such a GC threshold, we can well reconstruct the network topology of the I&F system.

For the nonlinear I&F dynamics, our preliminary investigation in numerical simulations shows that the order of lags needed in conditional probability is high and one suffers from the curse of dimensionality in an attempt to obtain good estimate of the transfer entropy. Therefore, how to characterize the directions of information flow using transfer entropy for the nonlinear I&F network dynamics remains another interesting and challenging issue.

ACKNOWLEDGMENTS

This work is supported by NYU Abu Dhabi Institute G1301 (Y.Z., Y.X., D.Z., D.C.), NSFC-91230202, Shanghai Rising-Star Program-15QA1402600 (D.Z.), NSF DMS-1009575 (D.C.), Shanghai 14JC1403800, 15JC1400104 and SJTU-UM Collaborative Research Program (D.C., D.Z.).

Appendix A: Unbiased PSD estimation

In this section, we demonstrate that $\mathbf{P}^m_{nu}(f)$ defined by Eq. (4) is an unbiased estimator of $\mathbf{P}^m(f)$. Without loss of generality, we consider the cross PSD $P^m_{nu,xy}$ of time series X_t and Y_t. The other PSD functions can be computed similarly as we replace X_t by Y_t or Y_t by X_t. By definition,

$$\begin{aligned} P^m_{nu,xy} &= \frac{NT}{N-1}\left[I^m_{nu,x}I^{m*}_{nu,y} - \frac{1}{N^2}\sum_{s=1}^{N} X^m_{t_s}Y^{m*}_{t_s}\right], \\ &= \frac{NT}{N-1}\left[\left(\frac{1}{N}\sum_{j=1}^{N} X^m_{t_j}e^{-\mathrm{i}2\pi f t_j}\right)\times \right. \\ &\quad \left.\left(\frac{1}{N}\sum_{k=1}^{N} Y^m_{t_k}e^{\mathrm{i}2\pi f t_k}\right) - \frac{1}{N^2}\sum_{s=1}^{N} X^m_{t_s}Y^{m*}_{t_s}\right], \\ &= \frac{T}{N\,(N-1)}\left[\sum_{j\neq k} X^m_{t_j}e^{-\mathrm{i}2\pi f t_j}Y^m_{t_k}e^{\mathrm{i}2\pi f t_k}\right]. \end{aligned}$$

Then, the expectation of $P^m_{nu,xy}$ with respect to different realizations of the N randomly chosen nonuniform sampling points for a fixed signal can be expressed as

$$\mathbb{E}\left(P^m_{nu,xy}\right) = \frac{T}{N\,(N-1)}\left[\sum_{j\neq k}\mathbb{E}\left(X^m_{t_j}e^{-\mathrm{i}2\pi f t_j}Y^m_{t_k}e^{\mathrm{i}2\pi f t_k}\right)\right].$$

Here, we have $\mathbb{E}\left(X^m_{t_j}e^{-\mathrm{i}2\pi f t_j}\right) = \frac{1}{T}\int_0^T X^m_t e^{-\mathrm{i}2\pi f t}\mathrm{d}t = I^m_x(f)$. For $j \neq k$, because t_j and t_k are chosen randomly and independently, $X^m_{t_j}e^{-\mathrm{i}2\pi f t_j}$ is independent of $Y^m_{t_k}e^{\mathrm{i}2\pi f t_k}$, which leads to $\mathbb{E}\left(X^m_{t_j}e^{-\mathrm{i}2\pi f t_j}Y^m_{t_k}e^{\mathrm{i}2\pi f t_k}\right) = \mathbb{E}\left(X^m_{t_j}e^{-\mathrm{i}2\pi f t_j}\right)\mathbb{E}\left(Y^m_{t_k}e^{\mathrm{i}2\pi f t_k}\right) = I^m_x(f)I^{m*}_y(f)$. Therefore, we obtain

$$\mathbb{E}\left(P^m_{nu,xy}\right) = TI^m_x(f)I^{m*}_y(f) = P^m_{xy}(f),$$

indicating that $P^m_{nu,xy}(f)$ is an unbiased estimator of $P^m_{xy}(f)$. Following the same computation, it is clear that $\mathbf{P}^m_{nu}(f)$ is an unbiased estimator of $\mathbf{P}^m(f)$.

Appendix B: Discontinuity of auto-covariance function

In this section, we present the computation of $\frac{\mathrm{d}C_{xx}(\tau)}{\mathrm{d}\tau}$ in detail to demonstrate its discontinuity at the origin for the I&F dynamics. By definition,

$$\left.\frac{\mathrm{d}C_{xx}(\tau)}{\mathrm{d}\tau}\right|_{0^+} = \lim_{\tau\to 0^+} -\mathbb{E}_t\left[\left(X(t)-m_x\right)X'(t-\tau)\right]. \tag{B1}$$

From the firing-reset dynamics of the I&F model (1), the jth firing event of neuron x at time $T_{x,j}$ leads to a discontinuity in the voltage trace $X(t)$. This jump discontinuity in turn results in a singularity in $X'(t)$, which can be described mathematically as $-(x^{\mathrm{th}}-x^{\mathrm{r}})\delta(t-T_{x,j})$, where $\delta(t)$ is the Dirac delta function. Therefore, we can separate $X'(t)$ as $X'(t) = X_c^{'}(t) - (x^{\mathrm{th}}-x^{\mathrm{r}})\sum_j \delta(t-T_{x,j})$, where $X_c^{'}(t)$ is the nonsingular part of $X'(t)$. Equation (B1) now reads

$$\left.\frac{\mathrm{d}C_{xx}(\tau)}{\mathrm{d}\tau}\right|_{0^+} = \lim_{\tau\to 0^+} -\mathbb{E}_t\Bigg[\left(X(t)-m_x\right)\Big(X_c^{'}(t-\tau) - (x^{\mathrm{th}}-x^{\mathrm{r}})\sum_j \delta(t-\tau-T_{x,j})\Big)\Bigg].$$

Note that $t = T_{x,j}+\tau$ as $\tau\to 0^+$ is the time right after the reset, therefore $X(t) = x^{\mathrm{r}}$ as $t\to T^+_{x,j}$ and $\lim_{\tau\to 0^+}\mathbb{E}_t\left[X(t)(x^{\mathrm{th}}-x^{\mathrm{r}})\sum_j\delta(t-\tau-T_{x,j})\right] = x^{\mathrm{r}}(x^{\mathrm{th}}-x^{\mathrm{r}})R_x$, where $\lim_{\tau\to 0^+}\mathbb{E}_t\left[\sum_j\delta(t-\tau-T_{x,j})\right] = \mathbb{E}_t\left[\sum_j\delta(t-T_{x,j})\right] = R_x$, R_x is the firing rate of neuron x. Then

$$\begin{aligned}
&\left.\frac{\mathrm{d}C_{xx}(\tau)}{\mathrm{d}\tau}\right|_{0^+}\\
&= -\lim_{\tau\to 0^+}\mathbb{E}_t\left[\left(X(t)-m_x\right)X_c^{'}(t-\tau)\right]\\
&\quad +(x^{\mathrm{r}}-m_x)(x^{\mathrm{th}}-x^{\mathrm{r}})R_x,\\
&= -\lim_{T\to+\infty,\tau\to 0^+}\frac{1}{T}\int_0^T\left(X(t+\tau)-m_x\right)X_c'(t)\mathrm{d}t\\
&\quad +(x^{\mathrm{r}}-m_x)(x^{\mathrm{th}}-x^{\mathrm{r}})R_x,\\
&= -\lim_{T\to+\infty}\frac{1}{T}\int_0^T\left(X(t)-m_x\right)\mathrm{d}X_c(t)\\
&\quad +(x^{\mathrm{r}}-m_x)(x^{\mathrm{th}}-x^{\mathrm{r}})R_x,
\end{aligned}$$

where $X_c(t) = \int_0^t X_c^{'}(s)\mathrm{d}s$ is a continuous function satisfying $X_c(T_{x,j+1}) - X_c(T_{x,j}) = x^{\mathrm{th}}-x^{\mathrm{r}}$. We separate the time T into interspike windows $(T_{x,j},T_{x,j+1}]$, $j=1,2,\cdots,N$, $T_{x,N+1}\leqslant T$ and $T_{x,N+2}\geqslant T$. It is easy to see that $X(t) = X_c(t) - X_c(T_{x,j}) + x^{\mathrm{r}}$ in an interspike time window $t\in(T_{x,j},T_{x,j+1})$, then we obtain

$$\begin{aligned}
&\left.\frac{\mathrm{d}C_{xx}(\tau)}{\mathrm{d}\tau}\right|_{0^+}\\
&= -\lim_{N\to+\infty}\frac{1}{T_{x,N+1}-T_{x,1}}\sum_{j=1}^N\Bigg[\int_{T_{x,j}}^{T_{x,j+1}}\left(X(t)-m_x\right)\\
&\quad \mathrm{d}\left(X_c(t)-X_c(T_{x,j})+x^{\mathrm{r}}\right)\Bigg] + (x^{\mathrm{r}}-m_x)(x^{\mathrm{th}}-x^{\mathrm{r}})R_x,\\
&= -\lim_{N\to+\infty}\frac{1}{T_{x,N+1}-T_{x,1}}\sum_{j=1}^N\left(\frac{1}{2}X^2\Big|_{x^{\mathrm{r}}}^{x^{\mathrm{th}}} - m_x\,X|_{x^{\mathrm{r}}}^{x^{\mathrm{th}}}\right)\\
&\quad +(x^{\mathrm{r}}-m_x)(x^{\mathrm{th}}-x^{\mathrm{r}})R_x,\\
&= -\lim_{N\to+\infty}\frac{N}{T_{x,N+1}-T_{x,1}}\frac{1}{2}\left(x^{\mathrm{th}}-x^{\mathrm{r}}\right)\left(x^{\mathrm{th}}+x^{\mathrm{r}}-2m_x\right)\\
&\quad +(x^{\mathrm{r}}-m_x)(x^{\mathrm{th}}-x^{\mathrm{r}})R_x.
\end{aligned}$$

As $\lim_{N\to+\infty}\frac{N}{T_{x,N+1}-T_{x,1}} = R_x$, we arrive at

$$\left.\frac{\mathrm{d}C_{xx}(\tau)}{\mathrm{d}\tau}\right|_{0^+} = -\frac{1}{2}\left(x^{\mathrm{th}}-x^{\mathrm{r}}\right)^2R_x,$$

which is Eq. (8) in the main text.

Appendix C: Discontinuity of cross-covariance function

In this section, we present the computation of $\frac{\mathrm{d}^2C_{xy}(\tau)}{\mathrm{d}\tau^2}$ and describe its discontinuity at the origin for the I&F dynamics. By definition, we obtain

$$\frac{\mathrm{d}^2C_{xy}(\tau)}{\mathrm{d}\tau^2} = -\frac{\mathrm{d}}{\mathrm{d}\tau}\mathbb{E}_t\left(X(t)Y'(t-\tau)\right).$$

Define $t' = t-\tau$, then

$$\begin{aligned}
\frac{\mathrm{d}^2C_{xy}(\tau)}{\mathrm{d}\tau^2} &= -\frac{\mathrm{d}}{\mathrm{d}\tau}\mathbb{E}_{t'}\left(X(t'+\tau)Y'(t')\right),\\
&= -\mathbb{E}_t\left(X'(t)Y'(t-\tau)\right),
\end{aligned}$$

which is Eq. (10) in the main text.

As discussed in Section IV C 2, we concentrate on $\mathbb{E}_t\left(S_T^x(t)S_T^y(t-\tau)\right)$ to understand the discontinuity behavior of $-\mathbb{E}_t\left(X'(t)Y'(t-\tau)\right)$ at the origin. Note that $\mathbb{E}_t\left(S_T^x(t)S_T^y(t-\tau)\right)$ is the covariance of the spike train

time series of neuron x and neuron y,

$$\begin{aligned}
&\mathbb{E}_t\left(S_T^x(t)S_T^y(t-\tau)\right)\\
&=\lim_{T\to+\infty}\frac{1}{T}\int_0^T S_T^x(t)S_T^y(t-\tau)\mathrm{d}t,\\
&=\lim_{T\to+\infty}\frac{1}{T}\int_0^T\left(\sum_{i=1}^{+\infty}\delta(t-T_{x,i})\right)\left(\sum_{j=1}^{+\infty}\delta(t-\tau-T_{y,j})\right)\mathrm{d}t,\\
&=\lim_{T\to+\infty}\frac{1}{T}\sum_{i,j,0<T_{x,i},T_{y,j}<T}\delta\left(\tau-(T_{x,i}-T_{y,j})\right),\\
&=R_{x|y}(\tau)R_y,
\end{aligned}$$

where $R_{x|y}(\tau)$ is the conditional firing rate of neuron x with time lag τ upon the firing of neuron y, *i.e.,* the spike-triggered firing rate of neuron x, R_y is the mean firing rate of neuron y. Without loss of generality, we consider a two-neuron I&F network with a unidirectional connection from neuron y to neuron x. We define $R^-_{x|y}=\lim_{\tau\to0^-}R_{x|y}(\tau)$, $R^+_{x|y}=\lim_{\tau\to0^+}R_{x|y}(\tau)$. In order to compute the difference between $R^-_{x|y}$ and $R^+_{x|y}$, we define $I^-_{x|y}(v)$ and $p^-_{x|y}(v)$ as the mean current flow and the stationary voltage probability density function, respectively, of neuron x at voltage v right before neuron y fires. Note that $R^-_{x|y}$ can be obtained as follows

$$R^-_{x|y}=p^-_{x|y}(x^{\mathrm{th}})I^-_{x|y}(x^{\mathrm{th}}).$$

Upon receiving a spike of neuron y, by the I&F dynamics (1), the current of neuron x will jump instantaneously. This current jump can be quantified by $\Delta I_x(v)=-s(v-x^{\mathrm{E}})$, where s is the coupling strength between neurons. Therefore,

$$\begin{aligned}
R^+_{x|y}&=p^-_{x|y}(x^{\mathrm{th}})\left[I^-_{x|y}(x^{\mathrm{th}})+\Delta I_x(x^{\mathrm{th}})\right],\\
&=R^-_{x|y}+s(x^{\mathrm{E}}-x^{\mathrm{th}})p^-_{x|y}(x^{\mathrm{th}}).
\end{aligned}$$

As $s(x^{\mathrm{E}}-x^{\mathrm{th}})p^-_{x|y}(x^{\mathrm{th}})>0$ for the I&F dynamics, we obtain

$$\begin{aligned}
&\mathbb{E}_t\left(S_T^x(t)S_T^y(t-\tau)\right)|_{\tau\to0^+}-\mathbb{E}_t\left(S_T^x(t)S_T^y(t-\tau)\right)|_{\tau\to0^-}\\
&=R^+_{x|y}(\tau)R_y-R^-_{x|y}(\tau)R_y,\\
&=s(x^{\mathrm{E}}-x^{\mathrm{th}})p^-_{x|y}(x^{\mathrm{th}})R_y,\\
&>0,
\end{aligned}$$

hence,

$$\begin{aligned}
&\left.\frac{\mathrm{d}^2C_{xy}(\tau)}{\mathrm{d}\tau^2}\right|_{0^+}-\left.\frac{\mathrm{d}^2C_{xy}(\tau)}{\mathrm{d}\tau^2}\right|_{0^-}\\
&\approx-s(x^{\mathrm{E}}-x^{\mathrm{th}})p^-_{x|y}(x^{\mathrm{th}})R_y\left(x^{\mathrm{th}}-x^{\mathrm{r}}\right)^2,\\
&<0,
\end{aligned}$$

which is the conclusion we present in the main text.

[1] C. W. J. Granger, Econometrica **37**, 424 (1969).
[2] J. Geweke, J. Amer. Statistical Assoc. **77**, 304 (1982).
[3] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, Phys. Lett. A **324**, 26 (2004).
[4] T. J. Mosedale, D. B. Stephenson, M. Collins, and T. C. Mills, J. Climate **19**, 1182 (2006).
[5] A. C. Lozano, N. Abe, Y. Liu, and S. Rosset, Bioinformatics **25**, i110 (2009).
[6] R. Nagarajan and M. Upreti, Stat. Appl. Genet. Mol. Biol. **9**, 1 (2010).
[7] C. Zou, C. Ladroue, S. Guo, and J. Feng, BMC Bioinformatics **11**, 337 (2010).
[8] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, Proc. Nat. Acad. Sci. U.S.A. **101**, 9849 (2004).
[9] A. Roebroeck, E. Formisano, and R. Goebel, Neuroimage **25**, 230 (2005).
[10] M. Ding, Y. Chen, and S. L. Bressler, in *Handbook of Time Series Analysis*, edited by B. Schelter, t. Winterhalder, and J. Timmer (Wiley-VCH Verlag GmbH & Co. KGaA, 2006) pp. 437–460.
[11] C. Hsiao, J. Econ. Dynam. Control **1**, 321 (1979).
[12] W. Joerding, J. Devel. Econ. **21**, 35 (1986).
[13] S. Z. Chiou-Wei, C.-F. Chen, and Z. Zhu, Energy Econ. Technological Change and the Environment, **30**, 3063 (2008).
[14] J. R. Freeman, Am. J. Polit. Sci. **27**, 327 (1983).
[15] L. Keele, Am. J. Polit. Sci. **51**, 241 (2007).
[16] J. T. Webb, M. A. Ferguson, J. A. Nielsen, and J. S. Anderson, PLoS ONE **8**, e84279 (2013).
[17] D. Zhou, Y. Xiao, Y. Zhang, Z. Xu, and D. Cai, Phys. Rev. Lett. **111**, 054102 (2013).
[18] D. Zhou, Y. Xiao, Y. Zhang, Z. Xu, and D. Cai, PLoS ONE **9**, e87636 (2014).
[19] E. Florin, J. Gross, J. Pfeifer, G. R. Fink, and L. Timmermann, J. Neurosci. Methods **198**, 344 (2011).
[20] D. Zhou, Y. Zhang, Y. Xiao, and D. Cai, New J. Phys. **16**, 043016 (2014).
[21] D. Zhou, Y. Zhang, Y. Xiao, and D. Cai, Front. Comput. Neurosci. **8**, 75 (2014).
[22] F. Marvasti, *Nonuniform Sampling: Theory and Practice* (Springer Science & Business Media, 2001).
[23] A. Aldroubi and K. Gröchenig, SIAM Rev. **43**, 585 (2001).
[24] P. Babu and P. Stoica, Digital Signal Process. **20**, 359 (2010).
[25] J. Scargle, Astrophys. J. **263**, 835 (1982).
[26] G. Foster, Astron. J. **112**, 1709 (1996).
[27] S. Vaughan, Philos. T. R. Soc. A **371**, 20110549 (2013).
[28] S. Baisch and G. H. R. Bokelmann, Comput. Geosci. **25**, 739 (1999).
[29] P. Zwartjes and M. Sacchi, Geophysics **72**, V21 (2006).
[30] F. J. Herrmann and G. Hennenfent, Geophys. J. Int. **173**, 233 (2008).

[31] M. Schulz and K. Stattegger, Comput. Geosci. **23**, 929 (1997).

[32] M. Schulz and M. Mudelsee, Comput. Geosci. **28**, 421 (2002).

[33] E. F. Glynn, J. Chen, and A. R. Mushegian, Bioinformatics **22**, 310 (2006).

[34] A. W. Liew, J. Xian, S. Wu, D. Smith, and H. Yan, BMC Bioinformatics **8**, 137 (2007).

[35] J. J. Benedetto and H.-C. Wu, Proc. SPIE **4119**, 130 (2000).

[36] M. Bourgeois, F. T. A. W. Wajer, D. v. Ormondt, and D. Graveron-Demilly, in *Modern Sampling Theory*, Applied and Numerical Harmonic Analysis, edited by J. J. Benedetto and P. J. S. G. Ferreira (Birkhäuser Boston, 2001) pp. 343–363.

[37] G. L. Bretthorst, in *AIP Conference Proceedings*, Vol. 567 (AIP Publishing, 2001) pp. 1–28.

[38] I. Bilinskis, *Digital Alias-free Signal Processing* (John Wiley & Sons, 2007).

[39] D. Rovnyak, D. P. Frueh, M. Sastry, Z.-Y. J. Sun, A. S. Stern, J. C. Hoch, and G. Wagner, J. Magn. Reson. **170**, 15 (2004).

[40] V. Jaravine, I. Ibraghimov, and V. Yu Orekhov, Nat. Methods **3**, 605 (2006).

[41] S. G. Hyberts, K. Takeuchi, and G. Wagner, J. Am. Chem. Soc. **132**, 2145 (2010).

[42] M. Dhamala, G. Rangarajan, and M. Ding, Neuroimage **41**, 354 (2008).

[43] R. Mirollo and S. Strogatz, SIAM J. Appl. Math. **50**, 1645 (1990).

[44] D. C. Somers, S. B. Nelson, and M. Sur, J. Neurosci. **15**, 5448 (1995).

[45] W. Mather, M. R. Bennett, J. Hasty, and L. S. Tsimring, Phys. Rev. Lett. **102**, 068105 (2009).

[46] Z. Wang, Y. Ma, F. Cheng, and L. Yang, Image Vis. Comput. **28**, 5 (2010).

[47] A. V. Rangan, D. Cai, and D. W. McLaughlin, Proc. Nat. Acad. Sci. U.S.A. **102**, 18793 (2005).

[48] D. Cai, A. V. Rangan, and D. W. McLaughlin, Proc. Nat. Acad. Sci. U.S.A. **102**, 5868 (2005).

[49] D. Zhou, A. V. Rangan, D. W. McLaughlin, and D. Cai, Proc. Nat. Acad. Sci. U.S.A. **110**, 9517 (2013).

[50] M. W. Maciejewski, H. Z. Qui, I. Rujan, M. Mobli, and J. C. Hoch, J. Magn. Reson. **199**, 88 (2009).

[51] L. Greengard and J. Lee, SIAM Rev. **46**, 443 (2004).

[52] G. T. Wilson, SIAM J. Appl. Math. **23**, 420 (1972).

[53] J. Geweke, J. Amer. Statistical Assoc. **79**, 907 (1984), 00305.

[54] G. Dumont, G. Northoff, A. Longtin, Phys. Rev. E **90**, 012702 (2014).

[55] P.E. Greenwood, M.D. McDonnell, L.M. Ward, Neural Comput. **27**, 74 (2015).

[56] L. Debnath and D. Bhatta, *Integral transforms and their applications* (CRC press, 2014).

[57] Y. Yuan, ICIAM **99**, 271 (2000).

[58] J. F. Epperson, Amer. Math. Monthly **94**, 329 (1987).

[59] E. Suli and D. F. Mayers, *An Introduction to Numerical Analysis* (Cambridge University Press, 2003).

[60] T. Schreiber Phys. Rev. Lett. **85**, 461 (2000).

[61] L. Barnett, A.B. Barrett, A.K. Seth, Phys. Rev. Lett. **103**, 238701 (2009).